

Example 1. Analysing visually recorded cover-abundance data

Andrea Onofri, Hans-Peter Piepho and Marcin Kozak

Contents

The dataset	1
A traditional ANOVA fit	1
A survival model	2
References	3

The dataset

The dataset refers to a field experiment aiming to compare the weed control abilities of nine post-emergence herbicides against *Sorghum halepense* in maize. Three weeks after the treatment, the cover-abundance of *S. halepense* was visually recorded in six classes, using the Braun-Blanquet method. The limits of the classes are shown as L (lower limit) and U (upper limit). The `midPoint` represents the centre of each class. Each dataset record represents one field plot.

First of all, we need to read the data into R. As the dataset is contained in the companion package ‘agriCensData’, we need to load this package first, assuming that it has been already installed in the system (as shown here). Together with this package, we also load all the other necessary packages.

```
library(agriCensData)
library(emmeans)
library(survival)
data(BBSurvey)
head(BBSurvey)
```

```
##   Plot Herbicide   L   U midPoint
## 1    1          A 0.1   5     2.55
## 2    2          A 0.1   5     2.55
## 3    3          A 5.0  25    15.00
## 4    4          A 5.0  25    15.00
## 5    5          B 0.1   5     2.55
## 6    6          B 0.1   5     2.55
```

A traditional ANOVA fit

Although we have actually collected the data by assigning each plot to one cover-abundance class, we may be tempted to use the mid-point of that class as the dependent variable. As this midpoint is a real number, we may fit a traditional ANOVA model. We can find the corresponding means using the `emmeans` package

(Lenth 2016). In the paper, we did not conduct pair-wise comparisons. If you want to do this anyway, you can do so using the same package, as shown below.

```
mod.aov <- lm(midPoint ~ Herbicide, data = BBSurvey)
means <- emmeans(mod.aov, ~ Herbicide)
means
```

```
## Herbicide emmean      SE df lower.CL upper.CL
## A          8.7750 5.501751 27 -2.513661 20.06366
## B          8.7750 5.501751 27 -2.513661 20.06366
## C          8.1375 5.501751 27 -3.151161 19.42616
## D          8.1500 5.501751 27 -3.138661 19.43866
## E          1.3000 5.501751 27 -9.988661 12.58866
## F         26.2500 5.501751 27 14.961339 37.53866
## G          8.7625 5.501751 27 -2.526161 20.05116
## H          8.1625 5.501751 27 -3.126161 19.45116
## I         56.2500 5.501751 27 44.961339 67.53866
##
## Confidence level used: 0.95
```

```
CLD(means, Letter = LETTERS, sort = F)
```

```
## Herbicide emmean      SE df lower.CL upper.CL .group
## A          8.7750 5.501751 27 -2.513661 20.06366 A
## B          8.7750 5.501751 27 -2.513661 20.06366 A
## C          8.1375 5.501751 27 -3.151161 19.42616 A
## D          8.1500 5.501751 27 -3.138661 19.43866 A
## E          1.3000 5.501751 27 -9.988661 12.58866 A
## F         26.2500 5.501751 27 14.961339 37.53866 A
## G          8.7625 5.501751 27 -2.526161 20.05116 A
## H          8.1625 5.501751 27 -3.126161 19.45116 A
## I         56.2500 5.501751 27 44.961339 67.53866 B
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```

Using this approach (the traditional ANOVA), we have circumvented the problem of censoring by *pretending* that the observations are more reliable than they actually are.

A survival model

The body of methods dealing with censored data is usually known as survival analysis, as data relating to the survival of individuals are very often censored. Obviously, we can use survival analysis with all types of censored data, also when they have nothing to do with the survival of individuals.

To fit a survival model, we need to load the survival package (Therneau 1999) and use the `survreg()` function from this package. Its arguments are similar to those used by the function `lm()`, with the only difference that the former method uses interval limits as the dependent variable, with no need for imputation. We argue that this is much more respectful of the manner the data were harvested with.

```
library(survival)
mod.surv <- survreg(
  Surv(L, U, type = "interval2") ~ Herbicide,
```

```

dist = "gaussian", data = BBsurvey)
means.surv <- emmeans(mod.surv, ~ Herbicide)
CLD(means.surv, Letters = LETTERS, sort = F)

```

```

## Herbicide      emmean      SE df  lower.CL  upper.CL .group
## A              6.755275 3.684287 26 -0.8178844 14.328434 AB
## B              6.755275 3.684287 26 -0.8178844 14.328434 AB
## C              5.903803 3.631445 26 -1.5607388 13.368345 A
## D              5.903803 3.631445 26 -1.5607388 13.368345 A
## E              1.269626 3.253208 26 -5.4174388  7.956691 A
## F             25.022493 4.008432 26 16.7830434 33.261942  B
## G              6.755275 3.684287 26 -0.8178844 14.328434 AB
## H              5.903803 3.631445 26 -1.5607388 13.368345 A
## I             57.335193 3.772500 26 49.5807077 65.089678  C
##
## Results are given on the Surv (not the response) scale.
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05

```

As we note in the main paper (*Literature reference, when available*), survival analysis estimates means with higher precision than the traditional ANOVA.

References

- LENTH, RV (2016) Least-Squares Means: The *R* Package **Lsmeans**. *Journal of Statistical Software* **69**
- THERNEAU, T (1999) *A Package for Survival Analysis in S*. R package version 2.36-14, Survival